

Differences of Pitch Profiles in Germanic and Slavic Languages

*Bistra Andreeva¹, Grażyna Demenko², Bernd Möbius¹,
Frank Zimmerer¹, Jeanin Jügler¹, Magdalena Oleskiewicz-Popiel²*

¹ Computational Linguistics & Phonetics, Saarland University, Germany

² Department of Linguistics, Adam Mickiewicz University, Poland

[andreeva, moebius, zimmerer, juegler]@coli.uni-saarland.de,
lin@amu.edu.pl, magdalena.jastrzebska@speechlabs.pl

Abstract

This study investigates cross-language differences in pitch range and variation in four languages from two language groups: English and German (Germanic) and Bulgarian and Polish (Slavic). The analysis is based on large multi-speaker corpora (48 speakers for Polish, 60 for each of the other three languages). Linear mixed models were computed that include various distributional measures of pitch level, span and variation, revealing characteristic differences across languages and between language groups. A classification experiment based on the relevant parameter measures (span, kurtosis and skewness values for pitch distributions for each speaker) succeeded in separating the language groups.

Index Terms: pitch range, pitch variation, cross-language differences, Bulgarian, Polish, German, British English

1. Introduction

Range and variation of fundamental frequency (f_0) are key ingredients of pitch profiles that have been shown to be characteristic for specific linguistic communities. Speakers of different languages, but also social groups within a single language, use a particular pitch profile as a distinguishing feature (see [8] for a review). For instance, Puerto Rican girls in New York City and age-matched native English speaking women use f_0 differently: Puerto Rican girls tend to speak on a higher pitch level than their native English peers [15]. Dialects of a language can also differ with respect to the use of f_0 (e.g. [7, 31]). Various cross-linguistic studies indicate language specific pitch profile differences. For instance, in a comparison of typologically different languages, English speakers were found to have a significantly lower median f_0 compared to speakers of Japanese, Spanish, and Tagalog [10, 11]. Other cross-language studies compared Polish vs. English [18], Mandarin vs. English [5, 12], British English vs. German [19, 20], or Russian vs. German [21]. Strong evidence for language-specific uses of pitch profiles has come from studies showing that bilingual speakers differ when speaking their two languages. For example, bilingual English/Japanese speakers used a higher pitch register in Japanese than in English [9, 30, 32]. Significant differences between language groups (Slavic vs. Germanic) have been reported in our previous study [1], showing that German and English speakers use lower pitch maxima, a narrower pitch span, and generally less variable pitch than Bulgarian and Polish speakers. Taken together, these findings demonstrate that differences in pitch profiles are not necessarily due only to possible physiological differences between speakers of different languages.

Complementing the analysis of speech *production* data, a number of studies have investigated the *perceptual* discrimination of languages based on their pitch profiles. Listeners can identify their own language based solely on prosodic cues, such as f_0 , amplitude, and timing ([22] for Cantonese, English, and Japanese). Even the f_0 contour alone has been shown to be a sufficient cue for discriminating pairs of languages, such as English and Japanese [24], English and French [16, 17], and English and Dutch [6]. Language specific profiles have also been found in the perception and production of paralinguistic attributes, such as politeness in Japanese vs. English [14] or the distinction between ‘confident’, ‘friendly’, ‘emphatic’ and ‘surprised’ in British English and Dutch [4].

Methodological differences make the comparison of findings across previous studies difficult. Such differences include, for instance, the discourse type and speaking style or the method of measuring f_0 contours. Moreover, many studies have been rather limited in terms of representativeness, being often based on a rather small number of speakers and sometimes restricted to either male or female speakers.

This paper extends our previous study [1] in two significant ways. First, the analyses presented here are based on a much larger pool of subjects, viz. 48 speakers of Polish and 60 speakers of each of the other three languages, Bulgarian, English and German, with an equal distribution of male and female speakers. This larger multi-speaker database significantly increases the robustness of statistical analyses. Second, we have performed a classification experiment based on the distributional measures that were found to be most characteristic for language-specific pitch profiles. The classification succeeded in providing a clear separation between the two language groups, Slavic and Germanic. Our findings support the hypothesis that linguistic communities tend to be characterized by particular pitch distribution profiles.

2. Material and Methods

Two Slavic (Bulgarian and Polish) and two Germanic (German and English) languages are under investigation in this study. The material analyzed is continuous read speech taken from two compatible, structurally similar multilingual speech databases, EUROM-1 (for German and English) [3] and BABEL (for Bulgarian and Polish) [26, 27]. We used a subset of the data, consisting of 3 cognitively linked short passages, containing 5 thematically connected sentences, read by 60 speakers (30 male and 30 female) for Bulgarian, German and English and 48 speakers (24 male and 24 female) for Polish. The passages were based on identical, real-life

topics for the four languages, freely translated and adapted for Bulgarian, German and Polish from the original English texts. The overall length of the analyzed material is about 70 minutes for Polish and 90 minutes for each of the other three languages.

3. f_0 Measures

Automatic pitch extraction was performed at intervals of 10 ms for male and 5 ms for female speakers using the RAPT algorithm [29] implemented in the program 'get_f0' from the ESPS software package. The f_0 tracks were checked and, if necessary, manually corrected. Speech samples with irregular voicing (e.g. creaky voice, laryngealization) were excluded from further analyses.

Pitch distribution profiles can be attributed to two related but distinct characteristics of a speaker's performance: (a) pitch level, i.e. the overall height of the speaker's voice, and (b) pitch span, i.e. the range of frequencies covered by the speaker [13]. To analyze cross-language differences in pitch range and variation, the following distributional measures were calculated per passage: mean and median f_0 values for level, and interquartile range (IQR) and the simple pitch excursion for span. The latter was computed as the difference between maximum and minimum pitch values per passage. The f_0 measurements for span were additionally converted to semitones by means of the formula (e.g. [25]):

$$39.863 * \log_{10}(\text{Maximum/Minimum})$$

The measures describing the variation and shape of the f_0 distribution were standard deviation (SD), kurtosis, and skewness (in Hz).

4. Results

4.1. General statistics

Means and standard deviations (in parentheses) of the distributional measures for pitch level and span are presented in Table 1, organized by language and speaker sex.

4.2. Linear mixed models

As a first step towards determining the differences, linear mixed models with the respective measure as dependent variable, speaker and passage as random factors, and native language (Bulgarian/Polish/English/German) and gender (male/female) as fixed factors, as well as all their possible interactions, were computed for each dependent variable in separate analyses by means of the JMP software [28]. Separate Tukey post-hoc tests were carried out per variable, if appropriate. The confidence level was set at $\alpha=0.05$.

Predictably, gender had a significant main effect on mean ($F [1, 220] = 1143.382, p<0.001$) and median f_0 ($F [1, 220] = 1032.336, p<0.001$), IQR ($F [1, 220] = 155.6597, p<0.001$), minimum f_0 ($F [1, 220] = 669.243, p<0.001$), maximum f_0 ($F [1, 220] = 807.7228, p<0.001$), f_0 span measured in Hz ($F [1, 220] = 270.4249, p<0.001$), SD ($F [1, 220] = 202.9187, p<0.001$) and skewness ($F [1, 220] = 7.8404, p<0.0056$), with females having significantly higher f_0 . Gender did not differ in kurtosis and f_0 span measured in semitones (see Figure 1 for f_0 mean, maximum, minimum and span measured in semitones). However, over and above the expected gender effect, there was also a significant main effect of language on all

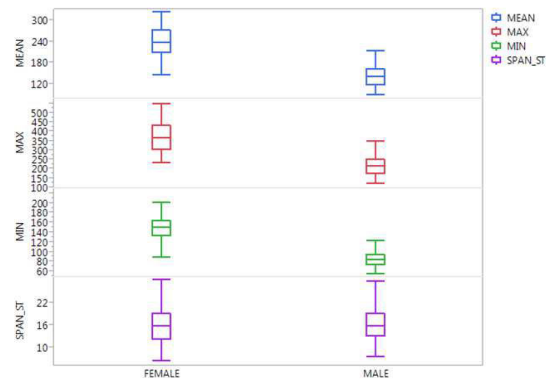


Figure 1: Gender main effect on f_0 mean, maximum, minimum and span in semitones

Table 1. Means and standard deviations of the distributional measures by language and speaker sex. Values in Hz except for the second span measure, which is in semitones (s.t.).

Measure	Bulgarian		Polish		German		English	
	male	female	male	female	male	female	male	female
mean	160 (21)	272 (32)	163 (22)	266 (24)	118 (16)	210 (20)	128 (22)	217 (20)
median	161 (22)	270 (34)	162 (24)	260 (24)	127 (22)	209 (21)	127 (22)	213 (21)
IQR	41 (13)	73 (20)	40 (12)	69 (22)	23 (7)	41 (12)	30 (15)	46 (17)
minimum	88 (15)	149 (25)	85 (15)	149 (21)	80.0 (12)	146 (25)	84 (13)	151 (23)
maximum	238 (37)	422 (52)	260 (37)	443 (62)	176 (29)	299 (31)	200 (43)	337 (53)
span	150 (37)	273 (49)	176 (36)	294 (66)	96 (26)	154 (35)	116 (39)	186 (61)
span (s.t.)	17.2 (3.6)	18.2 (3.0)	19.5 (3.4)	18.9 (3.7)	13.6 (2.7)	12.7 (3.5)	14.9 (3.4)	13.9 (4.1)
SD	29 (8)	52 (12)	32 (8)	53 (14)	17 (5)	28 (7)	22 (9)	35 (11)
skewness	.01 (.33)	.17 (.33)	.11 (.53)	.47 (.43)	.41 (.45)	.29 (.31)	.54 (.42)	.66 (.46)
kurtosis	-.26 (.48)	-.19 (0.46)	.29 (.54)	.28 (.86)	.30 (.93)	-.17 (.75)	.34 (1.00)	.51 (1.15)

measurements except on minimum f_0 , where the speakers are near the floor of their physiological f_0 range. Separate post-hoc tests showed that Bulgarian and Polish speakers had a significantly higher mean f_0 (F [3, 220] = 87.9677, $p < 0.001$), median f_0 (F [3, 220] = 83.4160, $p < 0.001$), IQR (F [3, 220] = 43.7389, $p < 0.001$) and f_0 span in semitones (F [3, 220] = 41.1905, $p < 0.001$) than English and German speakers. In the Slavic languages f_0 varies most strongly (possibly indicating more liveliness). Polish and Bulgarian reveal significantly higher SD values than English and German, although the English values are significantly greater than the German ones. The four languages differ significantly in their maximum f_0 values (F [3, 220] = 90.5398, $p < 0.001$). We found a positively skewed f_0 distribution for the four languages. This implies that the most frequent f_0 observation occurs lower than the mean. The skewness values for English speakers were significantly higher than those for German and Polish speakers and the values for the German and Polish speakers were significantly higher than those for Bulgarian speakers (F [3, 220] = 21.3182, $p < 0.001$). English speakers had a higher kurtosis than German and Bulgarian speakers, and Polish speakers had a higher kurtosis than Bulgarian speakers (F [3, 220] = 13.1106, $p < 0.001$). This reflects the fact that f_0 in Bulgarian and German is distributed over a narrower area (cf. Table 1).

The statistical analysis further revealed a significant interaction between language and gender for mean f_0 , median f_0 , IQR, maximum f_0 , SD, and skewness. This interaction can be explained by the higher f_0 register used by the Slavic speakers compared to the German speakers. The (relatively high) register for Polish and Bulgarian male speakers is in the same range of absolute f_0 values as that of English and German female speakers, causing them to group together in some analyses. Thus, the general pattern of higher f_0 values for the Slavic speakers than for Germanic speakers is retained (cf. Figure 2). These results are in line with our findings in [1]. Table 2 shows the f_0 measure patterns by languages.

Table 2. Language group differences for the f_0 measures on the basis of Tukey post-hoc comparisons.

f_0 measure	significant language-group differences
mean f_0	BG = PL > EN = DE
median f_0	BG = PL > EN = DE
min f_0	N.S.
IQR	BG = PL > EN = DE
max f_0	PL > BG > EN > DE
span s.t.	PL = BG > EN = DE
SD	PL = BG > EN > DE
skewness	EN > DE = PL > BG
kurtosis	EN = PL > PL = DE > BG

4.3. Classification with Multi-Layer Perceptrons

As discussed above, most f_0 measures - with the exception of min f_0 , which tends to reflect the lower physiological limit of f_0 production and is therefore quite stable across languages - appear to be characteristic of individual languages. However, the more general pattern that emerges from Table 2 is a separation of languages along the line of language group distinction, in that the Slavic languages (Bulgarian and Polish) as a group differ from the Germanic languages (English and German) for most f_0 measures in a consistent manner.

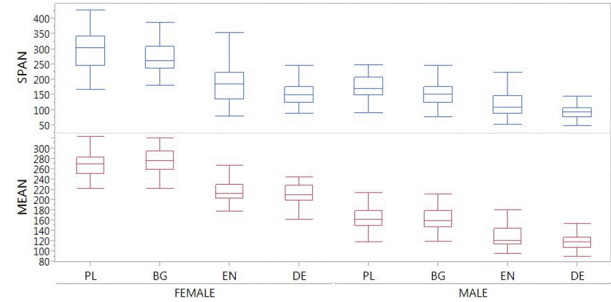


Figure 2: Mean and span f_0 values for Bulgarian, Polish, English and German speakers.

To estimate the strength of the contribution of f_0 measures to the language group distinction and at the same time add a method to distinguish datasets that are not linearly separable by classical statistical methods, a classification with Multi-Layer Perceptrons (MLP) with backpropagation learning was performed. The MLP net model was used as an attempt to explain the categorical variables (a) “language group” (Slavic vs. Germanic) and (b) “language” (Polish vs. German). The latter was included because the descriptive statistical analysis (variance analysis) showed the biggest differences for language pairs to exist between German and Polish.

Correlations between statistical distribution parameters were measured. The lowest correlations are between span, kurtosis, and skewness. Mean and median f_0 and IQR are significantly correlated with span.

An MLP with 3 input neurons equaling the number of input features (span, kurtosis, skewness), 7 hidden layers, and 2 output neurons for each language group was used because a performance maximum was observed using 7 neurons in the hidden layers compared to other net architectures (20 different nets were used for a preliminary evaluation of the quality of training). The outputs were normalized as posteriors by a softmax function. For the training 70% of the data were used, the validation and test sets each comprised 15% of the data.

Table 3. Results of training, validation, and testing.

	training		validation		testing	
	Germ.	Slav.	Germ.	Slav.	Germ.	Slav.
(a) Slavic vs. Germanic						
total #	250	230	51	51	59	43
correct	199	180	42	39	54	35
incorrect	51	50	9	12	5	8
% correct	79	78	82	76	91	81
% incorr.	20	21	17	23	8	18
(b) Polish vs. German						
	DE	PL	DE	PL	DE	PL
total #	122	106	29	19	29	19
correct	104	82	27	17	28	14
incorrect	18	24	2	2	1	5
% correct	85	77	93	89	96	73
% incorr.	14	22	6	10	3	26

Table 3 shows the results of the training, testing and validation based on three variables, f_0 span (in semitones), f_0 kurtosis and skewness. These three variables were selected because they are representative of pitch range and pitch variability, respectively, and no interactions were found between gender and language in the linear mixed models. Since the

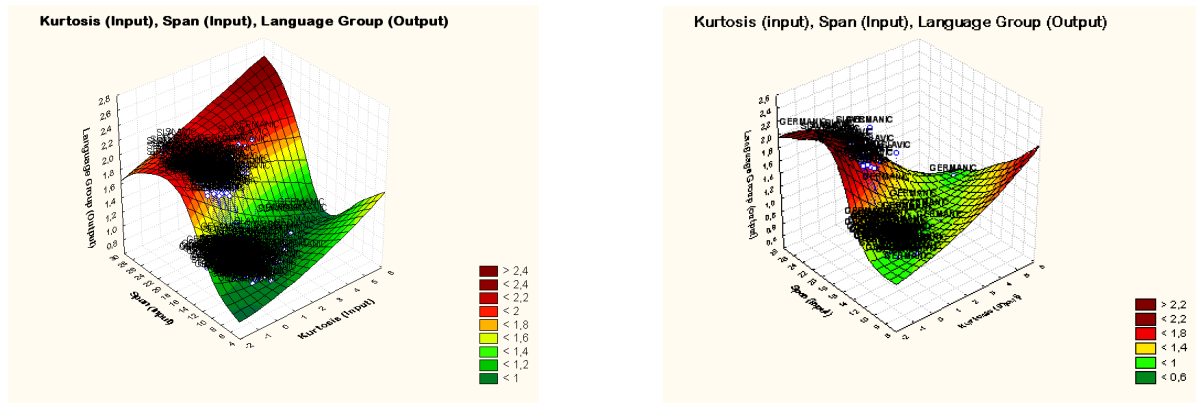


Figure 3: Language group classification (Slavic vs. Germanic) on the basis of f_0 span and kurtosis (left panel for training, right panel for testing)

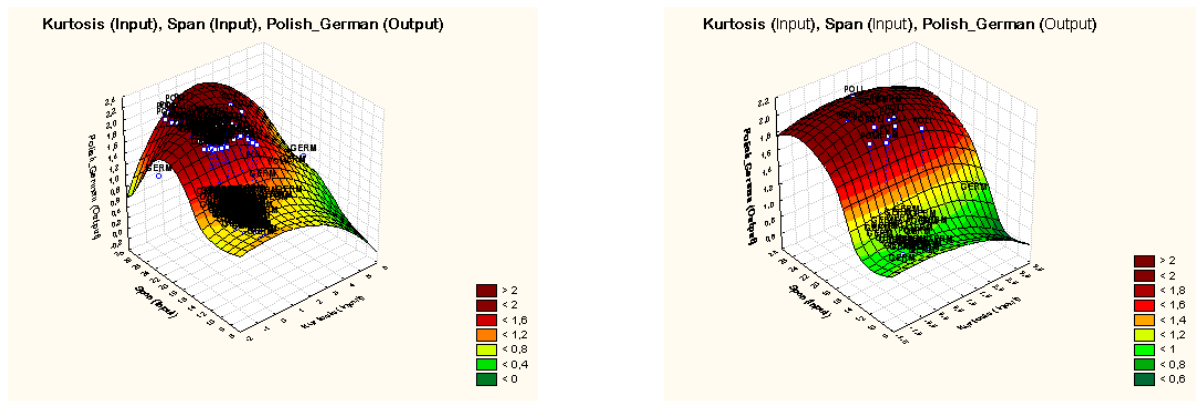


Figure 4: Language classification (Polish vs. German) on the basis of f_0 span and kurtosis (left panel for training, right panel for testing)

classification was carried out for male and female speakers combined, we expected these three variables to be key ingredients of language (group) specific pitch profiles.

Figures 3 and 4 provide a visual representation of f_0 span (in s.t.) and kurtosis (in Hz) for the Slavic and Germanic languages, and Polish and German, respectively. There is a clear separation between the groups – 91% correct classification for the Germanic and 81% for the Slavic language group; 96% for German and 73% for Polish (cf. Table 3). The Germanic/German speakers cluster in the lower right corner of the span/kurtosis plane, while the Bulgarian and Polish speakers cluster mostly in the higher left sector.

5. Discussion and Conclusions

This paper contributes to the growing number of studies on cross-language differences in pitch range and pitch variation. Our results are in line with our previous research [1] and confirm the hypothesis that linguistic communities tend to be characterized by particular pitch profiles. The male and female speakers of the Slavic group used considerably higher mean and median f_0 , IQR, span (s.t.) and maximum f_0 , and showed a larger SD than the speakers in the Germanic group. Classification with Multi-Layer Perceptrons with span, kurtosis and skewness as input variables showed a clear separation between the Germanic and Slavic group.

Further research using refined methods and measures of analysis is needed to explain this pattern. The measures used

in this study are global measures and therefore too general for a precise interpretation of the results. An alternative to measuring f_0 distribution is to reduce the f_0 contour to a series of target points representing the significant pitch changes by automatic stylization (cf. [2]) or by categorical pitch accent labeling. At this point, in terms of statistical pitch distribution patterns, we can only speculate that Slavic speakers may be generally more expressive in their use of f_0 features than Germanic speakers. Differences in pitch level and pitch span, however, may also reflect differences between languages in terms of tonal inventories and their phonetic implementation. Without access to an annotation of pitch accent types in our corpora, this is no more than an educated guess right now.

Therefore, in future work we will refine our measures of pitch range, by including linguistically based measures which were found to be better predictors of differences in pitch range and pitch variation across speakers and languages [2, 20, 23]. We also expect to add more languages (with using different factors for language categorization), as well as spontaneous speech data.

6. Acknowledgements

We wish to thank Ryszard Gubrynowicz (Institute of Fundamental Technology Research, Polish Academy of Science) and Snezhina Dimitrova (English Dept., Sofia University “St. Kliment Ohridski”), for kindly providing the Babel databases for Polish and Bulgarian, respectively.

7. References

- [1] Andreeva, B., Demenko, G., Wolska, M., Möbius, B., Zimmerer, F., Jügler, J., Jastrzebska, M., and Trouvain, J. (2014). Comparison of pitch range and pitch variation in Slavic and Germanic languages. *Proc. Speech Prosody* 2014, Dublin.
- [2] Campione, E., and Véronis, J. (1998). A statistical study of pitch target points in five languages. *Proceedings of ICSLP '98*, 1391-1394.
- [3] Chan, D. Fourcin, A.; Gibbon, D.; Granstrom, B.; Huckvale, M.; Kokkinakis, G.; Kvale, K.; Lamel, L.; Lindberg, B.; Moreno, A.; Mouropoulos, J.; Senia, F.; Trancoso, I.; Veld, C. and Zeiliger, J. (1995). Eurom - a spoken language resource for the EU. In *Eurospeech '95. Proceedings of the 4th European Conference on Speech Communication and Speech Technology*, 1, Madrid., 18-21 September 1995, 867-870.
- [4] Chen, A, Gussenhoven, C., and Rietveld, T. (2004). Language-specificity in the perception of paralinguistic intonational meaning, *Language & Speech* 47, 311-349.
- [5] Chen, G. T. (1972). *A comparative study of pitch range of native speakers of Midwestern English and Mandarin Chinese: An acoustic study*, doctoral dissertation, University of Wisconsin-Madison, Madison.
- [6] de Pijper, J. R. (1983). *Modelling British English intonation*, Dordrecht - Holland: Foris.
- [7] Deutsch, D., Le, J., Shen, J., and Henthorn, T. (2009). The pitch levels of female speech in two Chinese villages. *Journal of the Acoustical Society of America*, April, 125, EL208.
- [8] Dolson, M. (1994). The pitch of speech as a function of linguistic community, *Music Perception* 11 (3), 321-331.
- [9] Graham, C., (2013). Revisiting f0 Range Production in Japanese-English Simultaneous Bilinguals. *Annual Report of UC Berkeley Phonology Lab*, 110-125.
- [10] Hanley, T.D., Snidecor, J.C., and Ringel, R. (1966). Some acoustic differences among languages, *Phonetica* 14, 97-107.
- [11] Hanley, .D. and Snidecor, J.C. (1967). Some acoustic similarities among languages, *Phonetica* 17, 141-148.
- [12] Keating, P. & Kuo, G. (2012). Comparison of speaking fundamental frequency in English and Mandarin, *Journal of the Acoustical Society of America* 132, 1050-1060.
- [13] Ladd, D.R. (1996). *Intonational Phonology*. Cambridge: Cambridge University Press.
- [14] Loveday, L. (1981). Pitch, politeness and sexual role: an explanatory investigation into the pitch correlates of English and Japanese formula., *Language and Speech* 24, 71-89.
- [15] Luchsinger, R. and Arnold, G. (1965). *Voice-Speech-Language*. Constable&Co Ltd., London.
- [16] Maidment, J. A. (1976). Voice fundamental frequency characteristics as language differentiators. *Speech and hearing: Work in progress*, University College London, 74-93.
- [17] Maidment, J. A. (1983). Language recognition and prosody: further evidence, *Speech, hearing and language: Work in progress*, University College London 1, 133-141.
- [18] Majewski, W., Hollien, H., and Zalewski, J. (1972). Speaking fundamental frequency of Polish adult males, *Phonetica* 25, 119-125.
- [19] Mennen, I., Schaeffler, F., & Docherty, G. (2007). Pitching it differently: A comparison of the pitch ranges of German and English speakers. *16th International Congress of Phonetic Sciences (ICPhS XVI)*, Saarbrücken, 1769-1972.
- [20] Mennen, I., Schaeffler, F., & Docherty, G. (2012). Cross-language differences in fundamental frequency range: a comparison of English and German *Journal of the Acoustical Society of America* 131(3), 2249-2260.
- [21] Nebert, Augustin Ulrich (2013). *Der Tonhöhenumfang der deutschen und russischen Sprechstimme. Vergleichende Untersuchung zur Sprechstimmlage*. Hallesche Schriften zur Sprechwissenschaft und Phonetik, Band 46. Frankfurt/M.
- [22] Ohala, J. J., and Gilbert, J. B. (1979). Listeners' ability to identify languages by their prosody, in P. Léon and M. Rossi (eds.) *Problèmes de Prosodie*, Didier, Ottawa, 123-131.
- [23] Patterson, D. (2000). *A Linguistic Approach to Pitch Range Modelling*. Ph.D. dissertation, University of Edinburgh.
- [24] Ramus, F., and Mehler, J. (1999). Language identification with suprasegmental cues: A study based on speech resynthesis. *Journal of the Acoustical Society of America* 105 (1), 512-521.
- [25] Henning Reetz (1999): *Artikulatorische und akustische Phonetik*. Wissenschaftlicher Verlag, Trier.
- [26] Roach, P.; Arnfield, S. and Hallum, E., (1996). BABEL: A multi-language speech database. In *Proceedings of SST-96: Speech and Science Technology Conference*, Adelaide, 351-4.
- [27] Roach, P., Arnfield, S., Barry, W.J., Dimitrova, S., Boldea, M., Fourcin, A., Gonet, W., Gubrynowicz, R., Hallum, E., Lamel, L., Marasek, K., Marchal, A., Meister, E., Vicsi, K. (1998). Babel: a database of Central and Eastern European languages, *Proceedings of the First International Conference on Language Resources and Evaluation*, vol. 1, 28-30 May 1998, Granada, Spain, pp. 371-374.
- [28] SAS., "JMP" (Version 11), Cary (NC): SAS Institute, 2013.
- [29] Talkin, D. (1995). A Robust Algorithm for Pitch Tracking (RAPT). In Kleijn, W. B. and Paliwal, K. K. (eds.), *Speech Coding and Synthesis*. New York: Elsevier.
- [30] Todaka, Y. (1993). *A cross-language study of voice quality: bilingual Japanese and American speakers*, doctoral dissertation, University of California, Los Angeles, pp. 145-147.
- [31] Torgerson, R. C. (2005). *A comparison of Beijing and Taiwan Mandarin tone register: An acoustic analysis of three native speech styles*, master's thesis, Brigham Young University, 73-82.
- [32] Yamazawa, H., and Hollien, H. (1992). Speaking fundamental frequency pattern of Japanese women, *Phonetica* 49, 128-140.